

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-108844

(43)Date of publication of application : 12.04.2002

(51)Int.Cl.

G06F 17/21
G06F 12/00

(21)Application number : 2000-295352

(71)Applicant : HITACHI SOFTWARE ENG CO LTD

(22)Date of filing : 28.09.2000

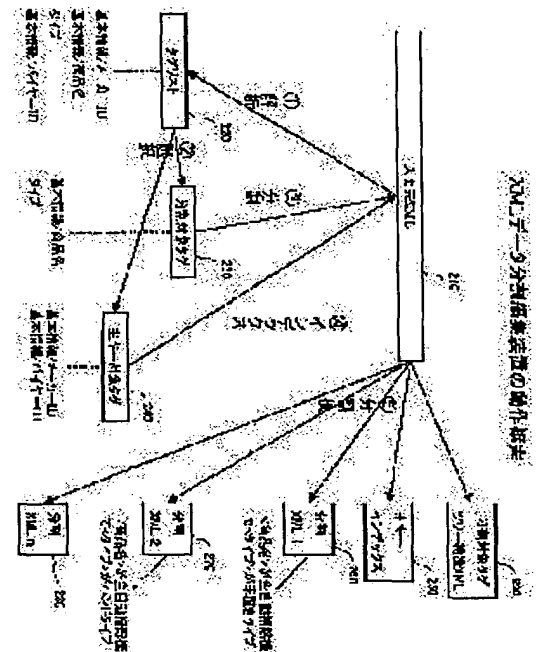
(72)Inventor : KIJIMA NORIKAZU

(54) XML DATA DIVISION EDITING DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To quickly retrieve and efficiently edit a large quantity of XML data by dividing them.

SOLUTION: Input source XML data 110 are analyzed to prepare a tag list 120. A main key index 150 for corresponding a main key to the split XML data, split XMLs 160, 170 and 180 divided by a split object tag and a split object tag tree structure XML 190 are prepared from the tag list 120 by using the split object tag 130, a main key object tag 140 and the data 110.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

タグ130と主キー対象タグ140を選択する。この分割対象タグツリー構造XML130と主キー対象タグ140は記憶装置60に格納される。処理装置110のXMLリーダー分割手段112は、入力元XMLデータ110を、分割対象タグ130によって分割して分割XMLデータ160、170、180を生成し、同時に、主キー対象タグ140の値と該当レコードが収められた分割XMLデータ160、170、180との関連付けを示し、主キーインデックスXML150、分割対象タグ130の値をタグの値として格納化した分割対象タグツリー構造XML190を生成し、外部記憶装置70に格納する。なお、タグリスト120なども外部記憶装置70に保持しておけば、あとで再活用が可能である。

【0010】以後、ユーザが検索要求や編集要求等があるとき、処理装置10では、主キーインデックス150をもとに分割対象タグツリー構造XML190を表示装置20に表示し、ユーザが分割対象タグを選択すると、該当分割XMLデータを外部記憶装置70から読み出しで表示装置20に表示する。

【0011】図7は入力元XML110の具体例（入力元XMLファイル710）を示す。入力元XMLファイル710は最初の商品データ720を含んでいる。商品データ720は「商品」タグのチルドに「基本情報」タグ721と「タイプ」タグ729を持つ。「基本情報」タグ721はチルドに「メーカーID」タグ722と「バイヤーID」タグ723、「商品名」タグ724、「消費電力」タグ725、「価格」タグ726を持つ。「価格」タグ726はチルドに「値」タグ727、「単位」タグ728を持つ。入力元XMLファイル710には、このような商品データが多数収められている。図8は主キーインデックスXML150の具体例（主キーインデックスXML810）を示す。図9は分割XML（1160）の具体例（分割XML910）を示す。図10は分割対象タグツリー構造XML190の具体例（分割対象タグツリー構造XML1010）を示す。

【0012】以下、図7乃至図10の具体例を参照しながら本XMLデータ分割編集装置のタグリスト生成手段11とXMLデータ分割手段12の動作を詳述する。

【0013】図3はタグリスト生成手段11の処理フローチャートである。このフローチャートに従って、商品データが収められたXMLファイル110からタグリストを生成する処理を以下に示す。

【0014】タグリスト生成手段11は、まず、一つの入力元XMLファイル710に着目する（ステップ30）。そして、1レコードを扱うタグがまず存在するか判定する（ステップ302）。図7に示す入力元XMLファイル710には商品名のデータが収められており、ルートタグ「document」の直下に「商品」タグがある。商品タグで囲まれた部分が一つの商品データを表

しており、1レコードとみなす。図7では、一つの商品（レコード）を扱うタグがまだ存在する。そこで、一つの商品データ（1レコード）720に着目し（ステップ303）、その中にまだ取り出していないタグがまだ存在するか判定する（ステップ304）。1商品データ720の中にはまだ取り出していないタグが存在する。そこで、一つのタグ「基本情報」721を取り出す（ステップ305）。そして、この取り出した値を保持し、分割対象タグに同名タグが存在しないか判定する（ステップ306）。「基本情報」タグ721は値を持たない。そこで、ステップ304に戻り、再び1商品データの中にまだ取り出していないタグが存在するか判定し、次のタグ「基本情報/メーカーID」722を取り出す（ステップ305）。「基本情報/メーカーID」タグ722は値を持ち、タグの同階層に同名タグを持たない。そこで、タグの名称「基本情報/メーカーID」722とタグの値「M0001」を取得し、「基本情報/メーカーID」722のタグの数を一つインクリメントし、「基本情報/メーカーID」タグ722の名称の数と1となる（ステップ307）。また、「基本情報/メーカーID」722のタグの値「M0001」の数を一つインクリメントし、「基本情報/メーカーID」タグ722の値の数と1となる（ステップ308）。

【0015】次に、ステップ304に戻り、「基本情報/バイヤーID」タグ723、「基本情報/商品名」タグ724、「基本情報/消費電力」タグ725、「基本情報/価格/値」タグ727、「基本情報/価格/単位」タグ728、「タイプ」タグ729の順に処理を繰り返し、一つの商品データ720の解析を終了する。

【0016】次に、ステップ302に戻り、同様に、入力元XMLファイル710にある商品の終わりまで繰り返す。この結果、入力元XMLファイル710の全商品（全レコード）について、1レコード中のタグ中でタグの同階層に一つのみ出現するタグの出現回数と同名タグで同じ値の数が全て解析される。

【0017】ここで、図7の入力元XMLファイル710の全商品データの解析結果は次のようになったとす

「タイプ」タグの数は100個、「タイプ」タグの値は「床置きタイプ」の数は60個、「タイプ」タグの値は「ハンンドタイプ」の数は40個ある。「基本情報/商品名」タグの数は100個、「基本情報/商品名」タグの値は「全自動掃除機」の数は40個、「基本情報/商品名」タグの値は「全自動掃除機」の数は35個、「基本情報/商品名」タグの値「全自動掃除機」の数は25個ある。「基本情報/消費電力」タグの値は「50W」のタグの数は65個、「基本情報/消費電力」タグの値は「300W」のタグの数は20個、「基本情報/消費電力」タグの値は「200W」のタグの数は15個ある。

【0018】図4はタグリストの具体例であり、図7の入力元XMLファイル710の全商品データの解析結果を上記のように仮定して、同名タグの数の多い順、同名タグで同じ値の数の多い順にタグ名とタグの値を並べたものである。図4中、410がタグリストを示す。タグリスト410は、入力XMLデータ710の中に含まれる1レコード（1商品データ）の中に一つだけ出現するタグ名412が入力XMLデータ710の中に含まれる同名タグ名413、同名タグで同じタグの値の数414の大きい順に順位411付けして並べたリストである。ここでは、同名タグで同じタグの値の数414は同名タグが同じ値をとる数を大きい順に数値目まで示し、タグリスト410では4番目まで出力されている。タグの数415は同名タグがとるタグの値の数値出力する。

【0019】例えば、入力XMLファイル710の中に含まれ、1レコードの中に一つだけ存在する「タイプ」タグの数は100個なので同名タグ数413は100個、入力XMLファイル710の中に含まれ、1レコードの中に一つだけ存在する「タイプ」タグの値は「床置きタイプ」が60個、「ハンンドタイプ」が40個あるので、同名タグで同じタグの値の数414は1番多いのが60個、2番目に多いのが40個となり、タグの数415には「床置きタイプ、ハンンドタイプ」と出力される。

【0020】ユーザは、タグリスト410から分割に適用切と思える分割対象タグ名420を選択する。また、タグリスト410から主キーとして使用する主キー対象タグ名430を選択する。ここでは、分割対象タグ名420として「基本情報/商品名」「タグ」と「タイプ」タグを選択する。主キー対象タグ名430としては「基本情報/メーカーID」と「基本情報/バイヤーID」タグを選択する。

【0021】図5はXMLデータ分割手段12の処理フローチャートである。以下、図5のフローチャートに従い、入力元XMLとして図7の入力元XML710を使い、分割対象タグ名として図4の「基本情報/商品名」タグと「タイプ」タグを指定し、主キー対象タグとして同じく図4の「基本情報/メーカーID」と「基本情報/バイヤーID」タグを指定した場合について説明する。

【0022】XMLデータ分割手段12は、まず、入力元XML710に着目する（ステップ501）。そして、まだ着目していないレコードは存在するか判定する（ステップ502）。入力元XML710では「商品」タグを1レコードとして取り扱う。入力元XML710にまだ着目していない商品データ720が存在する。そこで、入力元XML710において1レコード目の商品データ720を扱うタグ群に着目する（ステップ503）。そして、1レコード（1商品データ）の中にまだ着目していないタグは存在するか判定する（ステップ504）。1商品720を扱うタグ群の中でまだ着目してい

ないタグが存在する。「基本情報」タグ721に着目する（ステップ505）。そして、着目するタグは主キーのタグが判定する（ステップ506）。「基本情報」タグ721は主キーのタグではない。そこで、次に着目するタグは分割対象タグか判定する（ステップ508）。「基本情報」タグ721は分割対象タグではない。【0023】ステップ504に戻り、1レコード（1商品データ）の中にまだ着目していないタグは存在するか判定する。商品720を扱うタグ群の中で着目していないタグは存在する。「基本情報/メーカーID」タグ722に着目する。「基本情報/メーカーID」タグ722は主キーを扱うタグである（ステップ506）。「基本情報/メーカーID」タグ722の名前とタグの値「M0001」を保持する（ステップ507）。「基本情報/メーカーID」タグ722は分割対象タグではない（ステップ508）。

【0024】再びステップ504に戻る。商品710を扱うタグ群の中で着目していないタグは存在する。「基本情報/バイヤーID」タグ723に着目する（ステップ505）。「基本情報/バイヤーID」タグ723は主キーを扱うタグである（ステップ506）。「基本情報/バイヤーID」タグ723の名前とタグの値「B0001」を保持する（ステップ507）。「基本情報/バイヤーID」タグ723は分割対象タグではない（ステップ508）。

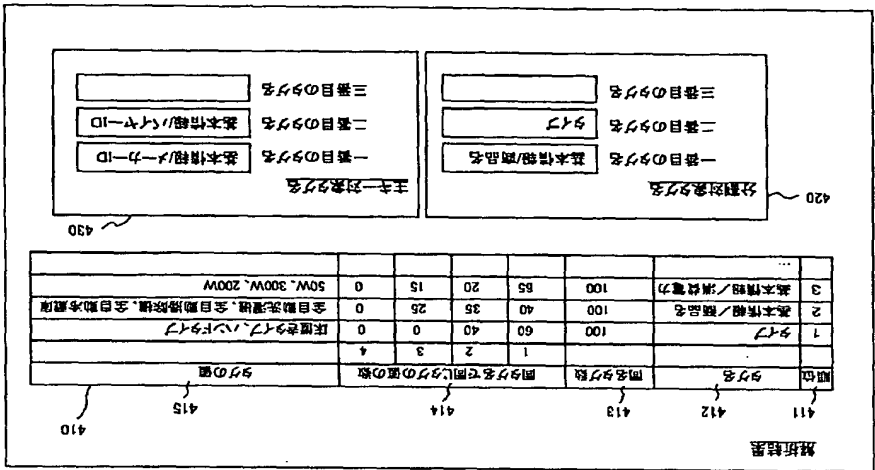
【0025】再びステップ504に戻る。商品720を扱うタグ群の中で着目していないタグは存在する。「基本情報/商品名」タグ724に着目する。「基本情報/商品名」タグ724は主キーを扱うタグではない（ステップ506）。「基本情報/商品名」タグ724は分割対象タグである（ステップ508）。そこで、「基本情報/商品名」タグ724の名前と値「全自動掃除機」の値を保持する（ステップ509）。

【0026】同様に、1商品データ720を扱うタグ群の残りのタグについても、ステップ504～509に基づいて処理する。残りのタグの中で主キー対象タグ、分割対象タグとして現れるものは、分割対象タグとして「タイプ」タグ729があり、「タイプ」タグ729の値は「床置きタイプ」である。

【0027】XMLデータ分割手段12は、1商品データを扱うタグ群の中で着目していないタグが存在しなくなった場合、主キー対象タグ名及びタグの値、並びに分割対象タグ名及びタグの値を主キーインデックスXML150に書き込む（ステップ510）。主キーインデックスXML150は主キー対象タグと格納先XMLとの関連付けを要す。

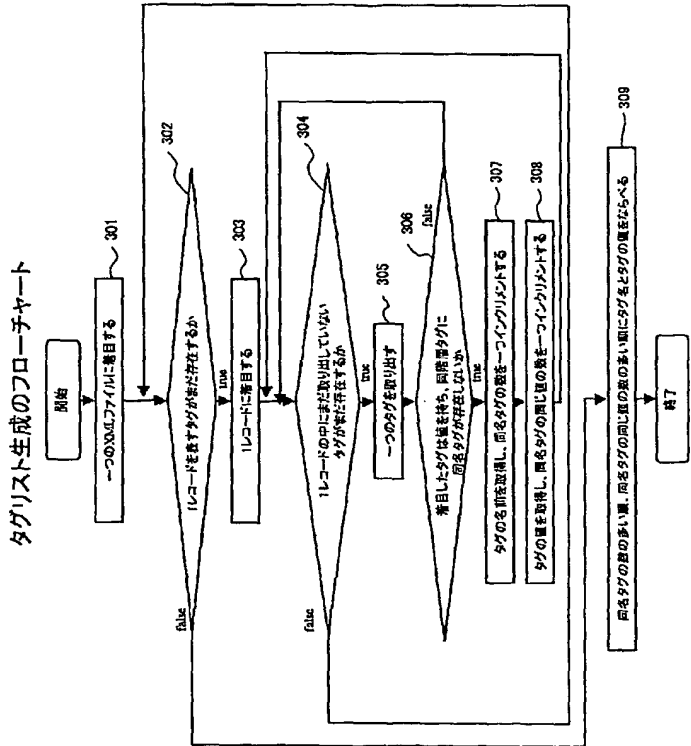
【0028】ここでは、主キーのタグ名、タグの値は「基本情報/メーカーID」タグ722とタグの値「M0001」、「基本情報/バイヤーID」タグ723とタグの値「B0001」である。分割対象タグ名、タグ

【図4】



タグリスト

【図3】



【図7】

入力元XMLデータの具体例

```
<?xml version="1.0" encoding="Shift_JIS"?>
<document>
  <商品>
    <基本情報>
      <メーカーID>M0001</メーカーID>
      <ハイヤーID>B0001</ハイヤーID>
      <商品名>全自動掃除機</商品名>
      <消費電力>550W</消費電力>
    </基本情報>
    <価格>
      <値>20000</値>
      <単位>円</単位>
    </価格>
    <基本情報>
      <タイプ>床置きタイプ</タイプ>
    </商品>
  </商品>
  <基本情報>
    <メーカーID>M0001</メーカーID>
    <ハイヤーID>B0021</ハイヤーID>
    <商品名>全自動掃除機</商品名>
    <消費電力>50W</消費電力>
  </商品>
  <基本情報>
    <タイプ>ハンドタイプ</タイプ>
  </商品>
  ...
</document>
```

主キーインデックスXMLの具体例

【図8】

```
<?xml version="1.0" encoding="Shift_JIS"?>
<document>
  <商品>
    <主キー対象>
      <メーカーID>M0001</メーカーID>
      <ハイヤーID>B0001</ハイヤーID>
    </主キー対象>
    <基本情報>
      <商品名>全自動掃除機</商品名>
    </基本情報>
    <タイプ>床置きタイプ</タイプ>
  </商品>
  <商品>
    ...
  </商品>
  <商品>
    ...
  </商品>
  ...
</document>
```

【図10】

分割対象タグツリー構造XMLの具体例

```
<?xml version="1.0" encoding="Shift_JIS"?>
<document>
  <全自動洗濯機>
    <床置きタイプ>
    <全自動洗濯機>
    <全自動冷蔵庫>
    <床置きタイプ>
    <全自動冷蔵庫>
    <全自動掃除機>
    <床置きタイプ>
    <ハンドタイプ>
    <全自動掃除機>
  </document>
```

【図9】

分割後XML 1の具体例

910

```
<?xml version="1.0" encoding="Shift_JIS"?>
<document>
  <商品>
    <基本情報>
      <メーカーID>M0001</メーカーID>
      <バイヤーID>B0001</バイヤーID>
      <商品名>全自動掃除機</商品名>
      <消費電力>350W</消費電力>
    </基本情報>
    <価格>
      <値>20000</値>
      <単位>円</単位>
    </価格>
    <基本情報>
      <タイプ>床置きタイプ</タイプ>
    </商品>
    ...
  </document>
```